Data meet Theories: Up Close and Personal

Ioannis Votsis
University of Düsseldorf
votsis@phil-fak.uni-duesseldorf.de

Jim Bogen and James Woodward's 'Saving the Phenomena', published only twenty years ago, has become a modern classic. Their centrepiece idea is a distinction between data and phenomena. According to them, data are typically the kind of things that are observable or measurable like "bubble chamber photographs, patterns of discharge in electronic particle detectors and records of reaction times and error rates in various psychological experiments" (p. 306). Phenomena are physical processes that are typically unobservable. Examples of the latter category include "weak neutral currents, the decay of the proton, and chunking and recency effects in human memory" (ibid.). Theories, in Bogen and Woodward's view, are utilised to systematically explain and predict phenomena, not data (pp. 305-306). The relationship between theories and data is rather indirect. Data count as evidence for phenomena and the latter in turn count as evidence for theories. This view has been further elaborated in subsequent papers (see Bogen and Woodward 1992, 2005 and Woodward 1989) and is becoming increasingly influential (e.g. Prajit K. Basu 2003, Stathis Psillos 2004 and Mauricio Suárez 2005). In this paper I argue that in various significant and well-known cases theories accompanied with suitable auxiliary hypotheses are more proximal to observations than Bogen and Woodward would have us believe. This is especially true of cases involving novel predictions.

The first two case studies I intend to discuss come from Bogen and Woodward's original article. One concerns the melting point of lead. According to Bogen and Woodward's analysis, the relevant data in this case are scatter points of temperature readings. These cannot be explained by, derived or predicted from theoretical considerations. Such considerations relate only to the relevant phenomena – in this case the phenomenon that lead melts at 327.5 °C. In their view, we can say that the mean of the observed distribution is a good estimate for the true melting point of lead (and hence evidence for the phenomenon). We cannot however explain, infer or predict the given data points from the theory plus any suitable auxiliary hypotheses because the mean does not: (i) represent a property of, or necessarily coincide with, any single data point and (ii) coincide exactly with the true value.

I argue against this line of reasoning by emphasising two points. First, we need not infer particular data points but rather confidence intervals. Second, the same considerations employed to estimate the mean can be used to construct auxiliary hypotheses that together with the main theory facilitate the inference, prediction and potentially the explanation of relevant features of the data. One of the auxiliaries in this case looks something like this: If the temperature of substance X is between 327.4 °C and 327.6 °C, the instrument registers values between 327.4 and 327.6 respectively.

The other Bogen and Woodward case concerns the detection of weak neutral currents, a phenomenon predicted by the Weinberg-Salam electroweak model. The relevant data in this case are bubble chamber photographs. In their attempt to detect weak neutral currents, scientists at the Gargamelle bubble chamber in CERN ran into the following difficulty. Neutrons were thought to be confounding factors as their collision with nucleons produced the same signature as weak neutral currents – which were the product of neutrino interactions with nucleons. The obvious proposal was to calculate the contribution of these neutrons and then somehow to deduct it from the total effect. The perhaps not so obvious problem was that estimating such contributions was far from easy. Two solutions were put forth: (a) to run monte carlo simulations in order to establish the upper bound of neutron-induced effects and

(b) on the assumption that neutron-induced effects are more likely to occur along the bubble chamber's wall, to check by empirical measurement whether this is reflected in the spatial distribution of putative weak neutral currents. Bogen and Woodward maintain that in both solutions the information linking data and phenomena was neither sufficiently detailed nor certain and for this reason it cannot be utilised to reconstruct an explanation, prediction or inference from the theory to the data.

As with the melting point of lead case, the presence or absence of suitable auxiliaries is crucial. Either the information linking data with phenomena is unreliable and hence does not confer confirmation of the relevant phenomena or it is reliable and hence can be used to construct auxiliary hypotheses that supplement the Weinberg-Salam model thereby allowing us to infer, predict and potentially explain the data from largely theoretical considerations. In solution (a) one of the relevant auxiliaries looks like this: If our quasi-theoretical estimates of the frequency of weak neutral currents are reliable, branching tracks with geometrical properties y should be detectable with frequency x. In solution (b), and as Bogen and Woodward inadvertently concede, "theoretical reasons" were used to calculate that "neutron-induced events would occur more frequently near the walls of the chamber" (1988, p.330). One of the relevant auxiliaries thus looked like this: If neutron-induced events occur more frequently along the periphery, the spatial distribution of branching tracks with geometrical properties y should be non-uniform between the centre of a photograph and its periphery.

Last, yet far from least, I consider a number of paradigmatic novel prediction cases that lend themselves to the same analysis. Considered by many as the Holy Grail in confirmation, novel predictions are difficult to achieve unless data and theories are intimately related. Take the prediction of what we now call the 'Poisson spot'. Siméon-Denis Poisson, a particle theorist vis-à-vis light, derived from Augustin Fresnel's wave theory of light the following unexpected observational prediction: A bright spot should appear in the middle of a disk's circular shadow when illuminated by a narrow beam of light. François Arago performed the experiment and to everyone's disbelief the bright spot was observed. As a consequence Fresnel's wave theory received a hard-earned confirmational boost. Without the auxiliary hypothesis that *constructive interference implies brighter regions*, Poisson would not have been able to predict the bright spot. This and other cases make it painfully obvious that suitable auxiliaries (connecting theories and data) are often, if not always, available and inform scientists about the observable manifestations of physical phenomena. As a result even well entrenched theories can be undone when the right data comes along.

References:

Basu, P. K. (2003) 'Theory-ladenness of Evidence: A Case Study from History of Chemistry', *Studies in the History and Philosophy of Science Part A*, vol. 34: 351-368.

Bogen, J., and J. Woodward (1988) 'Saving the phenomena', *The Philosophical Review*, vol. 97: 303–352.

Bogen, J., and J. Woodward (1992) 'Observations, Theories and the Evolutions of the Human Spirit', *Philosophy of Science*, vol. 59(4): 590-611.

Bogen, J., and J. Woodward (2005) 'Evading the IRS', *Poznan Studies in the Philosophy of the Sciences and the Humanities, Idealization XII*: *Correcting the Model*, Martin R. Jones and Nancy Cartwright (eds.), Amsterdam: Rodopi, pp. 233-268.

Psillos, S. (2004) 'Tracking the Real: Through Thick and Thin', *British Journal for the Philosophy of Science*, vol. 55, 393-409.

Suarez, M. (2005) 'The Semantic View, Empirical Adequacy, and Application', *Crítica Revista Hispanoamericana de Filosofía*, vol. 37, no. 109, 29-63.

Woodward, J. (1989) 'Data and Phenomena', *Synthese*, vol. 79(3): 393-472.